DOCUMENT RESUME

ED 419 008 TM 028 305

AUTHOR Vos, Hans J.
TITLE Applications of Bayesian Decision Theory to Sequential Mastery Testing.
INSTITUTION Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.
REPORT NO RR-97-06
PUB DATE 1997-00-00
NOTE 36p.
AVAILABLE FROM Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Bayesian Statistics; Concept Formation; Cutting Scores; Foreign Countries; Higher Education; *Mastery Tests; Medical Education; *Psychometrics; Test Construction
IDENTIFIERS *Decision Theory; *Sequential Testing

ABSTRACT

The purpose of this paper is to formulate optimal sequential rules for mastery tests. The framework for this approach is derived from empirical Bayesian decision theory. Both a threshold and linear loss structure are considered. The binomial probability distribution is adopted as the psychometric model involved. Conditions sufficient for sequentially setting optimal cutting scores are presented. Optimal sequential rules will be derived for the case of a beta distribution representing prior true level functioning. An empirical example of sequential mastery testing for concept-learning in medicine concludes the paper. (Contains 6 tables and 51 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 419 008

TM

Applications of Bayesian Decision Theory to Sequential Mastery Testing

Research Report 97-06

Hans J. Vos

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- ☑ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- • Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

IM028305

BEST COPY AVAILABLE

Applications of Bayesian Decision Theory to Sequential Mastery Testing

Hans J. Vos

Abstract

The purpose of this paper is to formulate optimal sequential rules for mastery tests. The framework for the approach is derived from empirical Bayesian decision theory. Both a threshold and linear loss structure are considered. The binomial probability distribution is adopted as the psychometric model involved. Conditions sufficient for sequentially setting optimal cutting scores are presented. Optimal sequential rules will be derived for the case of a beta distribution representing prior true level of functioning. An empirical example of sequential mastery testing for concept-learning in medicine concludes the paper.

Applications of Bayesian Decision Theory to Sequential Mastery Testing

In a fixed-length mastery test, the decision is to classify students as either a master or a nonmaster. During the last two decades, the fixed-length mastery problem has been studied extensively by many researchers (e.g., Cronbach & Gleser, 1965; Davis et al., 1973; De Gruijter & Hambleton, 1984; Hambleton & Novick, 1973; Huynh, 1976, 1980; Swaminathan et al., 1975; van der Linden, 1980, 1990; van der Linden & Mellenbergh, 1977; Wilcox, 1977). Most of these authors derived, analytically or numerically, optimal rules by applying (empirical) Bayesian decision theory (e.g., DeGroot, 1970; Lehmann, 1959; Lindgren, 1976) to this problem. The application of (empirical) Bayesian methods to decision making consists of two basic elements: A psychometric model relating observed test scores and student's true level of functioning to each other, and a loss structure evaluating the total costs and benefits of all possible decision outcomes. Optimal rules are derived by minimizing the posterior expected loss.

Beside the fixed-length mastery problem, attention has also been paid to the variable-length mastery problem. In this type of problem the decision is to classify students as a master, a nonmaster, or present another item. The main goal of a variable-length mastery test is to provide shorter tests for students who have clearly attained a certain level of mastery (or clearly nonmastery) and longer tests for those students for whom the mastery decision is not as clear-cut (Lewis & Sheehan, 1990). In case the items are randomly selected, the variable-length mastery problem is also known as a sequential or multistage mastery problem. If a computer is used for administering and scoring the items (e.g., Lewis & Sheehan, 1990; Sheehan & Lewis, 1992), and the optimal rule is determined using sequential decision theory, the mastery test is called a computerized mastery test (CMT).

One of the earliest sequential mastery tests was designed by Ferguson (1969a, 1969b) using Wald's sequential probability ratio test (SPRT). In Ferguson's approach, students' responses to items are assumed to follow a binomial probability distribution. The binomial model assumes that, given the true level of functioning, the probability to answer the item correctly is equal for all items in the pool, or that items are sampled at random. Using item response theory (IRT) models, Reckase (1983) and Kingsbury and Weiss (1983) proposed alternative sequential mastery testing procedures within an SPRT-framework. In both procedures, as opposed to Ferguson's approach, items are not assumed to have equal difficulty but are allowed to vary in difficulty and discrimination. In addition, the next item to be

BEST COPY AVAILABLE

presented to the student is not selected randomly but is based on the principle of maximizing the amount of information. Hence, the item selection procedures proposed by Reckase (1983) and Kingsbury and Weiss (1983) are adaptive instead of random (also see Spray & Reckase, 1996).

In the Lewis and Sheehan (1990) model, Bayesian theory is used to determine the optimal sequential number of equivalent testlets (i.e., short blocks of parallel items) to be randomly administered to the student. As in Reckase (1983) and Kingsbury and Weiss (1983), the conditional probability of a correct response, given the true level of functioning, is modeled using IRT. A threshold loss function is assumed from which the posterior expected losses associated with the mastery and nonmastery decisions can be calculated at each stage of sampling. The posterior expected loss associated with continuing sampling is determined considering all possible decision outcomes of future randomly presented items by backward induction. The optimal sequential decision rule is now found by selecting the action (i.e., mastery, nonmastery, or to continue sampling) that minimizes posterior expected loss at each stage of sampling. Doing so, as indicated by Lewis and Sheehan (1990), the action selected at each stage of sampling is optimal with respect to the entire sequential mastery testing procedure.

The purpose of the present paper is to derive optimal rules for sequential mastery tests. As in the Lewis and Sheehan model, optimal sequential rules are determined using Bayesian decision theory. Our approach differs from Lewis and Sheehan, however, in the following five respects. First, as in Ferguson's approach, for the conditional probability of a correct response given the true level of functioning (i.e., the psychometric model), the binomial instead of an IRT model is considered. Two, in addition to threshold loss, optimal sequential rules are also derived for linear loss. Three, conditions sufficient for sequentially setting optimal cutting scores are presented. Four, optimal sequential rules will be derived when prior true level of functioning is determined through an analysis of empirical data (i.e., empirical Bayesian approach) instead of through a subjective assessment. It will be assumed in the present paper that prior true level of functioning can be characterized by a beta distribution, which will be examined against the data in the empirical example. Finally, four instead of three possible actions are distinguished, namely the action administering one more randomly selected item and the three classification actions mastery, partial mastery, and nonmastery. The paper concludes with an empirical example of a computerized mastery test for concept-learning in medicine.

The Sequential Four-Action Mastery Problem

In the following, a sequential four-action mastery test is supposed to have a maximum length of n ($n \geq 1$). Following Ferguson (1969a, 1969b), a maximum test length is specified for those students for whom it is very difficult to classify them as a master, partial master, or nonmaster. Let the observed item response at each stage of sampling k ($1 \leq k \leq n$) for a randomly sampled student be denoted by a discrete random variable X_k, with realization x_k. The observed response variable X_k takes the value 0 for a correct response and 1 for an incorrect response to the kth item. The variables $X_1,...,X_k$ are assumed to be independent and identically distributed for each value of k ($1 \leq k \leq n$). Let $S_k = X_1 +...+ X_k$ ($1 \leq k \leq n$) be the observed number-correct score variable, with realization $s_k = x_1 +...+ x_k$. ($0 \leq s_k \leq k$). Furthermore, due to measurement and sampling error, the sequential four-action mastery test is assumed not to be a perfect indicator of student's true performance. Therefore, let student's true level of functioning $t \in [0,1]$ at each stage of sampling k ($1 \leq k \leq n$) be denoted by a continuous random variable T.

Suppose $X_1 = x_1,...,X_k = x_k$ has been observed. Then the two fundamental elements of the application of Bayesian methods to sequential decision making discussed earlier can be formulated as follows: A loss function describing the loss $l(a_i(x_1,...,x_k),t)$ incurred when action $a_i(x_1,...,x_k)$ is taken for the student whose true level of functioning is t, and a psychometric model relating observed number-correct score s_k to student's true level of functioning t at each stage of sampling k ($1 \leq k \leq n$). In fact, it is the unreliability of the test that opens the possibility of applying (sequential) Bayesian methods to the problem of determining the optimal number of items (Hambleton & Novick, 1973).

In the sequential four-action mastery problem, given $X_1 = x_1,...,X_k = x_k$, the following four actions are available to the decision-maker at each stage of sampling k ($1 \leq k < n$): First, declare nonmastery to a student, $a_1(x_1,...,x_k)$, if his/her number-correct score s_k is equal to or below a certain cutting score $s_{c1}(k)$ on the observed number-correct score scale S_k. Second, declare partial mastery to a student, $a_2(x_1,...,x_k)$, if his/her number-correct score s_k exceeds $s_{c1}(k)$ but is below a certain cutting score $s_{c2}(k)$ on S_k, where $s_{c1}(k) < s_{c2}(k)$. Three, declare mastery to a student, $a_3(x_1,...,x_k)$, if his/her number-correct score s_k is equal to or exceeds $s_{c2}(k)$. Fourth, continue sampling, $a_4(x_1,...,x_k)$, if the posterior expected loss associated with administering one more random item is minimal. For the final stage of sampling, n, only the

three classification actions nonmastery, partial mastery, and mastery are available to the decision-maker.

It is important to notice that, linking up with common practice in criterion-referenced testing, the optimal sequential rules w.r.t. the three mastery classification decisions are assumed to have monotone forms; that is, rules in the form of cutting scores $s_{c1}(k)$ and $s_{c2}(k)$. Conditions sufficient for optimal sequential rules to be monotone are given later on.

Let the criteria levels t_{c1} and t_{c2} ($0 \leq t_{c1} < t_{c2} \leq 1$) represent the highest and lowest true level of functioning at which a student will be considered a true nonmaster and a true master, respectively. Furthermore, a student will be considered a partial true master if his/her true level of functioning exceeds t_{c1} but is below t_{c2}. The two criteria levels t_{c1} and t_{c2} must be specified in advance by the decision-maker (e.g., Angoff, 1971; Ebel, 1972; Nedelsky, 1954). Given the values of the criteria levels t_{c1} and t_{c2} on T, the sequential four-action mastery problem can now be stated at each stage of sampling k ($1 \leq k < n$) as choosing values of $s_{c1}(k)$ and $s_{c2}(k)$ or continue sampling such that the posterior expected loss is minimal. For the final stage of sampling, n, our sequential mastery problem reduces to choosing values of $s_{c1}(n)$ and $s_{c2}(n)$ such that the posterior expected loss is minimal.

Loss Structure

Generally speaking, a loss function evaluates the total costs and benefits of all possible decision outcomes for a student whose true level of functioning is t. These costs may concern all relevant psychological, social, and economic consequences which the decision brings along (e.g., extra computer time associated with presenting randomly additional items). The Bayesian approach allows the decision-maker to incorporate into the decision process the costs of misclassifications (i.e., students for whom the wrong decision is made).

In this section, as in the Lewis and Sheehan model, first the well-known threshold loss function will be discussed. Next, it is argued that in many situations the linear loss structure is a more realistic representation of the losses actually incurred.

Threshold Loss

The choice of this function implies that the costs and benefits involved can be summarized by possibly different constants for each possible decision outcome. Although this function may be less realistic in some applications, it has been studied extensively in the psychometric

literature, in particularly in the (sequential) mastery testing literature (e.g., Ben-Shakhar & Beller, 1983; Chuang et al., 1981; Davis et al. 1973; Hambleton & Novick, 1973; Huynh, 1976; Lewis & Sheehan, 1990; Novick & Lewis, 1974; Raju et al, 1991; Swaminathan et al., 1975).

Following Lewis and Sheehan (1990), a threshold loss function for our sequential mastery problem can be formulated as a natural extension of the one for the standard fixed-length two-action problem at each stage of sampling k ($1 \leq k \leq n$) as follows:

Table 1. Table for threshold loss function at stage k ($1 \leq k \leq n$) of sampling.

True Level / Action	$T \leq t_{c1}$	$t_{c1} < T < t_{c2}$	$T \geq t_{c1}$
$a_1(x_1, ..., x_k)$	ke	l_{12} + ke	l_{13} + ke
$a_2(x_1, ..., x_k)$	l_{21} + ke	ke	l_{23} + ke
$a_3(x_1, ..., x_k)$	l_{31} + ke	l_{32} + ke	ke

Just as in the Lewis and Sheehan model, the value e represents the costs of administering one random item. For the sake of simplicity, again following Lewis and Sheehan, the costs of administering one random item are assumed to be equal for each decision outcome as well as for each sampling occasion. Of course, these two assumptions can be relaxed in specific sequential mastery testing applications.

When optimizing the decision rule, a loss function needs to be determined only up to a positive multiplicative constant and an additive constant (e.g., Luce & Raiffa, 1957). Therefore, assuming the losses l_{11}, l_{22}, and l_{33} associated with the correct decision outcomes are equal and take the smallest values, the threshold loss function in Table 1 was rescaled in such a way that l_{11}, l_{22}, and l_{33} were equal to zero. Consequently, the rescaled losses l_{ij} (i,j = 1,2,3; $i \neq j$) associated with the incorrect decisions must take positive values.

Furthermore, it follows immediately from the way actions $a_i(x_1,...,x_k)$ (i = 1,2,3) were defined that action $a_1(x_1,...,x_k)$ is most appropriate when t is small, whereas action $a_2(x_1,...,x_k)$ is most appropriate when t takes intermediate values, and action $a_3(x_1,...,x_k)$ is most appropriate when t is large. As a result, the loss functions associated with actions $a_1(x_1,...,x_k)$ and $a_3(x_1,...,x_k)$ must be nondecreasing and nonincreasing in t, respectively. As far as the loss function associated with action $a_2(x_1,...,x_k)$ (i.e., partial mastery) is concerned, it cannot be

BEST COPY AVAILABLE

determined beforehand whether the loss for a student whose true level of functioning is below t_{c1} (i.e., l_{21}) is equal to, larger than, or smaller than the loss for a student whose true level of functioning exceeds t_{c2} (i.e., l_{23}). We only know that the loss associated with the correct partial mastery decision, l_{22}, must be smallest.

The loss parameters l_{ij} ($i = 1,2,3$; $i \neq j$) have to be empirically assessed. For assessing loss functions empirically, most texts on decision theory propose lottery methods (e.g., Luce & Raiffa, 1957, Chap.2). In general, these methods use the notions of desirability of outcomes to scale the consequences of each pair of actions and true level of functioning. It may be noted that, in addition to lottery methods, other psychological scaling methods can be used for assessing empirically loss parameters as well. For instance, van der Gaag, Mellenbergh, and van den Brink (1988), van der Gaag (1990), and Vrijhof, Mellenbergh, and van den Brink (1983) empirically assessed loss functions using Bechtel's preference method (Bechtel, 1976) and Comrey's constant sum method (Torgerson, 1958).

Linear Loss

An obvious disadvantage of the threshold loss function is that it assumes that, for instance, the same constant loss holds for all 'masters' whose true level of functioning is to the right of t_{c2}, no matter how large their distance from t_{c2}. It seems more realistic to suppose that for true masters the loss is a monotonically decreasing function of t (van der Linden, 1980).

Moreover, the threshold loss function is discontinuous; at the criteria levels t_{c1} and t_{c2} this function "jumps" from one constant value to another. This sudden change seems unrealistic in many real-life decision making situations. In the neighborhood of these points, the losses for correct and incorrect decisions should change smoothly rather than abruptly (Davis et al., 1973).

To overcome these shortcomings, van der Linden and Mellenbergh (1977) proposed a continuous loss function for the fixed-length two-action mastery problem which is a linear function of student's true level of functioning t (see also van der Linden & Vos, 1996; Vos, 1990, 1991, 1995, 1997, 1998). For our sequential mastery problem, their linear loss function can be restated at each stage of sampling k ($1 \leq k \leq n$) as follows (see also Davis et al., 1973):

$$l(a_i(x_1,\ldots,x_k),t) = \begin{cases} b_1(t - t_{c1}) + ke & \text{for } i = 1 \\ b_2(t - t_{c2}) + ke & \text{for } i = 2 \\ b_3(t_{c2} - t) + ke & \text{for } i = 3, \end{cases} \qquad (1)$$

where b_i, $(b_1-b_2) > 0$ (i = 1,2,3).

At each stage of sampling k ($1 \leq k \leq n$), the above defined function consists for each action $a_i(x_1,....,x_k)$ (i = 1,2,3) of a constant term and a term proportional to the difference between the true level of functioning t and the specified criterion level t_{c1} or t_{c2}. Analogous to the threshold loss function, the constant amounts of loss, e, associated with administering one random item are assumed to be equal for each action as well as for each sampling occasion. The condition $b_1, b_2, b_3 > 0$ is equivalent to the statement that for action $a_1(x_1,....,x_k)$ and $a_2(x_1,....,x_k)$, loss is assumed to be a strictly increasing function of t whereas loss for action $a_3(x_1,....,x_k)$ is assumed to be strictly decreasing in t. Furthermore, the condition $(b_1-b_2) > 0$ states that the loss for action $a_1(x_1,....,x_k)$ increases more quickly in t than for action $a_2(x_1,....,x_k)$.

It should be noted that the linear loss function seems to be a realistic representation of the losses actually incurred in many decision making situations. In a recent empirical study, van der Gaag (1990) showed that for various real-life fixed-length mastery decisions in psychology and education the loss structures can be approximated satisfactory by linear functions.

The loss parameters b_i (i = 1,2,3) have to be assessed empirically again.

Binomial Distribution as a Psychometric Model

To determine the optimal sequential number of items, a psychometric model to specify the statistical relation between the observed number-correct score and student's true level of functioning at each stage of sampling is needed. In the present paper, following Ferguson (1969a, 1969b), the well-known binomial model will be adopted.

As indicated by van den Brink (1982), when tests are assumed sampled from item domains, as in our sequential four-action mastery problem, the well-known binomial model is a natural choice for estimating the distribution of student's number-correct score s_k and making classification decisions (mastery, partial mastery, nonmastery). The binomial model assumes that the probability function relating the observed number-correct score s_k ($0 \leq s_k \leq k$) to student's true level of functioning t, $f(s_k|t)$, at stage k can be written as follows:

$$f(s_k|t) = \binom{k}{s_k} t^{s_k} (1-t)^{k-s_k}. \qquad (2)$$

Conditions Sufficient for Sequentially Setting Cutting Scores

As far as the three mastery classification decisions concern, as noted earlier, we confine ourselves in this paper to monotone rules. The restriction to monotone rules, however, is correct only if it can be proven that for any nonmonotone rule for the problem at hand there is a monotone rule with at least the same value on the criterion of optimality used (Ferguson, 1967, p.55). In a Bayesian fashion, the posterior expected loss is taken as the criterion of optimality.

The posterior expected loss for continuing sampling is determined by averaging the posterior expected loss associated with each of the possible future decision outcomes relative to the probability of observing those outcomes (Lewis & Sheehan, 1990). Therefore, it follows immediately that the conditions sufficient for setting cutting scores for the fixed-length three-action mastery problem at each stage of sampling, are also sufficient for the sequential four-action mastery problem. Generally, conditions sufficient for setting cutting scores for the fixed-length multiple-decision problem are given in Ferguson (1967, p.286).

First, the probability model relating observed number-correct score s_k to student's true level of functioning t, $f(s_k|t)$, must have a monotone likelihood ratio (MLR); that is, it is required that for any $t_1 > t_2$, the likelihood ratio $f(s_k|t_1)/f(s_k|t_2)$ is a nondecreasing function of s_k. MLR implies that a high true level of functioning tends to coincide with a high observed number-correct score. Second, the condition of monotone loss must hold; that is, there must be an ordering of the actions such that for each pair of adjacent actions the loss functions have at most one point in which the difference between the losses changes sign.

The condition of MLR holds for the binomial distribution, since this distribution belongs to the one-parameter exponential family which is well known to have MLR (e.g., Hogg & Craig, 1978). Generally, as shown by Gray (1988), for $f(s_k|t)$ to have MLR it is sufficient to show that the items have nondecreasing item characteristic functions.

Assuming the indices reflect the proper ordening of the actions, it follows from Table 1 that for threshold loss the condition of monotone loss is satisfied if at each stage of sampling k ($1 \le k \le n$):

$$(l_{13}+ke) - (l_{23}+ke) \ge (l_{12}+ke) - ke \ge ke - (l_{21}+ke),$$
$$(l_{23}+ke) - ke \ge ke - (l_{32}+ke) \ge (l_{21}+ke) - (l_{31}+ke). \quad (3)$$

or, equivalently,

$$l_{13} - l_{23} \geq l_{12} \geq - l_{21},$$
$$l_{23} \geq - l_{32} \geq l_{21} - l_{31}. \quad (4)$$

Since (1) implies that $[l(a_1(x_1,...,x_k),t)-l(a_2(x_1,...,x_k),t)] = [(b_1-b_2)t-b_1t_{c1}+b_2t_{c2}]$ and $[l(a_2(x_1,...,x_k),t)-l(a_3(x_1,...,x_k),t)] = [(b_2+b_3)(t-t_{c2})]$, it follows immediately from b_i, $(b_1-b_2) > 0$ that the condition of monotone loss is also satisfied for linear loss at each stage of sampling k $(1 \leq k \leq n)$.

Optimal Rules for the Sequential Four-Action Mastery Problem

In this section, optimal cutting scores will be derived for the sequential four-action mastery problem. Doing so, first the posterior expected loss for the fixed-length three-action mastery problem will be minimized, given $X_1 = x_1,...,X_k = x_k$ $(1 \leq k \leq n)$. In other words, for the fixed-length three-action mastery problem it will be determined which of the three actions $a_1(x_1,...,x_k)$, $a_2(x_1,...,x_k)$, or $a_3(x_1,...,x_k)$ yields the smallest posterior expected loss, given an observed item response vector $(x_1,...,x_k)$. Next, optimal rules for the sequential four-action mastery problem are computed at each stage of sampling k $(1 \leq k \leq n)$ by comparing this smallest posterior expected with the posterior expected loss associated with action $a_4(x_1,...,x_k)$ (i.e., continuing sampling).

Minimal Posterior Expected Loss for the Fixed-Length Mastery Problem

In minimizing the posterior expected loss for the fixed-length three-action mastery problem, first the situation with linear loss will be elaborated. Next, the case of threshold loss will be examined. It will be assumed that the empirical data fits a beta distribution, which is used to represent prior knowledge about T.

Appropriate Mastery Classification Decision with Linear Loss

It can easily be verified from (1) that the decision rule minimizing the posterior expected loss in the case of linear loss, given $X_1 = x_1,...,X_k = x_k$ $(1 \leq k \leq n)$, is to declare mastery, $a_3(x_1,...,x_k)$, when a student's number-correct score s_k $(0 \leq s_k \leq k)$ is such that

$$E[b_3(t_{c2}-T)+ke \mid s_k] \leq E[b_2(T-t_{c2})+ke \mid s_k]. \quad (5)$$

Since $(b_2+b_3) > 0$, this is equivalent to declare mastery if

$$E(T \mid s_k) \geq t_{c2}, \quad (6)$$

where $E(T \mid s_k)$ denotes the posterior expectation of T, given the observed number-correct score s_k.

If the inequality in (6) does not hold, a decision rule minimizing the posterior expected loss, given $X_1 = x_1,...,X_k = x_k$ $(1 \leq k \leq n)$, is to declare partial mastery, $a_2(x_1,...,x_k)$, if it holds for number-correct score s_k that

$$E[b_2(T-t_{c2})+ke \mid s_k] \leq E[b_1(T-t_{c1})+ke \mid s_k], \quad (7)$$

and to declare nonmastery $(a_1(x_1,...,x_k))$ to him/her otherwise. Since $(b_1-b_2) > 0$, it follows that partial mastery is declared if

$$E(T \mid s_k) \geq (b_1t_{c1}-b_2t_{c2})/(b_1-b_2), \quad (8)$$

and nonmastery is declared otherwise.

Putting $l(a_2(x_1,...,x_k),t)$ and $l(a_3(x_1,...,x_k),t)$ equal to each other, it appears that the t-coordinate of the intersection of both loss lines, t_{c2}, is equal to the right-hand side of (6). Similarly, the t-coordinate of the intersection of $l(a_1(x_1,...,x_k),t)$ and $l(a_2(x_1,...,x_k),t)$, say t_{12}, is equal to the right-hand side of (8).

Hence, with linear loss, the decision procedure for the fixed-length three-action mastery problem, given $X_1 = x_1,...,X_k = x_k$ $(1 \leq k \leq n)$, can now be stated as follows: Mastery is declared to a student $(a_3(x_1,...,x_k))$ if his/her posterior expectation of T is equal to or larger

than t_{c2}. If his/her posterior expectation of T is smaller than t_{c2}, however, the following two situations can be distinguished: First, his/her posterior expectation of T is smaller than t_{c2} but equal to or larger than t_{12}. In this case, partial mastery is declared ($a_2(x_1,...,x_k)$). Second, not only his/her posterior expectation of T is smaller than t_{c2} but also smaller than t_{12}. In this case, nonmastery is declared ($a_1(x_1,...,x_k)$).

In the present paper, prior knowledge about T will be estimated by using empirical data from other students of the group to which the individual student belongs (i.e., empirical Bayes approach). Here, it will be assumed that the empirical data fits a beta distribution, $B(\alpha,\beta)$. Its flexible form nearly always makes an approximation of prior beliefs possible (Novick & Jackson, 1974, p. 107-113). In the empirical example, it will be examined if this assumption holds against the data.

Keats and Lord (1962) have shown that simple moment estimators of α and β, based upon the mean μ and the KR-21 reliability ρ of the observed number-correct score from other students of the group to which the student belongs, are given as

$$\hat{\alpha} = (-1+1/\rho)\mu,$$
$$\hat{\beta} = -\hat{\alpha} + m/\rho - m \qquad (9)$$

where m denotes the number of items in the test from which ρ and μ are computed.

It follows from an application of Bayes' theorem that under the assumed binomial model from (2), the posterior distribution of T will again be a member of the beta family (the conjugacy property, see e.g., Lehmann, 1959). In fact, if the prior distribution is $B(\alpha,\beta)$ and student's observed number-correct score is s_k from a test of length k ($1 \leq k \leq n$), then the posterior distribution is $B(\alpha+s_k,\beta+k-s_k)$.

Using the fact that the expectation of a beta distribution $B(\alpha,\beta)$ is equal to $\alpha/(\alpha+\beta)$, it follows that the posterior expectation of T can be written as $(\alpha+s_k)/(\alpha+s_k+\beta+k-s_k)$, or, equivalently,

$$E(T \mid s_k) = (\alpha+s_k)/(\alpha+\beta+k). \qquad (10)$$

Hence, with linear loss and a beta distribution representing prior knowledge about T, the optimal number of items for the fixed-length three-action mastery problem, given X_i =

$x_1,...,X_k = x_k$ $(1 \leq k \leq n)$, can be computed by comparing the right-hand side of (10) with the right-hand sides of (6) and (8).

As an aside, it may be noted that if no information is available from the group to which the individual student belongs, the parameters of the beta prior can be specified as $\alpha = \beta = 1$. In that case, the prior distribution represents a uniform distribution on the standard interval from zero to one; hence, prior true level of functioning can take on all values between 0 and 1 with equal probability.

It is important to notice that if no empirical data is available for estimating prior true level of functioning, we are no longer dealing with an empirical Bayesian approach. Prior knowledge about T is estimated in this case by subjective assessment (e.g., Lewis & Sheehan, 1990).

Appropriate Mastery Classification Decision with Threshold Loss

In the case of threshold loss, it can be seen from Table 1 that a decision rule minimizing the posterior expected loss, given $X_1 = x_1,...,X_k = x_k$ $(1 \leq k \leq n)$, is to declare mastery $(a_3(x_1,...,x_k))$ when a student's number-correct score s_k $(0 \leq s_k \leq k)$ is such that

$$l_{31}P(T \leq t_{c1} \mid s_k) + l_{32}P(t_{c1} < T < t_{c2} \mid s_k) + ke \leq$$
$$l_{21}P(T \leq t_{c1} \mid s_k) + l_{23}P(T \geq t_{c2} \mid s_k) + ke. \qquad \textbf{(11)}$$

Rearranging terms, it can easily be verified from (11) that mastery is declared if

$$(l_{31}-l_{21}-l_{32})P(T \geq t_{c1} \mid s_k) + (l_{23}+l_{32})P(T \geq t_{c2} \mid s_k) - l_{31}+l_{21} \geq 0. \qquad \textbf{(12)}$$

If the inequality in (12) does not hold, a decision rule minimizing the posterior expected loss, given $X_1 = x_1,...,X_k = x_k$, is to declare partial mastery, $a_2(x_1,...,x_k)$, when a student's number-correct score s_k $(0 \leq s_k \leq k)$ is such that

$$l_{21}P(T \leq t_{c1} \mid s_k) + l_{23}P(T \geq t_{c2} \mid s_k) + ke \leq$$
$$l_{12}P(t_{c1} < T < t_{c2} \mid s_k) + l_{13}P(T \geq t_{c2} \mid s_k) + ke, \qquad \textbf{(13)}$$

and to declare nonmastery $(a_1(x_1,...,x_k))$ otherwise. It follows that partial mastery is declared if

BEST COPY AVAILABLE

$$(l_{12}+l_{21})P(T \geq t_{c1} \mid s_k) + (l_{13}-l_{23}-l_{12})P(T \geq t_{c2} \mid s_k) - l_{21} \geq 0, \quad \textbf{(14)}$$

and nonmastery is declared otherwise.

The cumulative posterior distributions $P(T \geq t_{c1} \mid s_k)$ and $P(T \geq t_{c2} \mid s_k)$ in (12) and (14) of the beta prior have been extensively tabulated (e.g., Pearson, 1930). Normal approximations are also available (Johnson & Kotz, 1970, sect. 2.4.6). In general, if T has a beta distribution with parameters (α,β) where neither α nor β is small (say, not < 10), then this distribution can be approximated by a normal distribution with mean $\alpha/(\alpha+\beta)$ and variance $\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$.

Minimizing Posterior Expected Loss for the Sequential Mastery Problem

Since the action $a_4(x_1,...,x_n)$ (i.e., continuing sampling) is not available at the final stage of sampling, n, the action $a_1(x_1,...,x_n)$, $a_2(x_1,...,x_n)$, or $a_3(x_1,...,x_n)$ with the smallest posterior expected loss also represents the optimal sequential rule at the final stage of sampling. Optimal sequential rules at the other stages of sampling k (i.e., $1 \leq k < n$) are computed by comparing the smallest posterior expected loss of the three actions $a_1(x_1,...,x_k)$, $a_2(x_1,...,x_k)$, and $a_3(x_1,...,x_k)$ with the posterior expected loss of action $a_4(x_1,...,x_k)$. As noted before, the posterior expected loss associated with continuing sampling is determined by considering all possible future decision outcomes (i.e., backward induction). Hence, the following backward induction computational scheme can be used for determining the optimal sequential rules for our four-action mastery problem:

Suppose that $X_1 = x_1,...,X_n = x_n$ has been observed at the final stage of sampling, n. Then, it is first computed which of the three actions $a_1(x_1,...,x_n)$, $a_2(x_1,...,x_n)$, and $a_3(x_1,...,x_n)$ yields the smallest posterior expected loss at the final stage of sampling. Let this optimal action be denoted as $\varphi_n(x_1,...,x_n)$ and its associated minimum posterior expected loss as $V_n(x_1,...,x_n)$.

Generally, the action $a_1(x_1,...,x_k)$, $a_2(x_1,...,x_k)$, or $a_3(x_1,...,x_k)$ yielding the smallest posterior expected loss, given $X_1 = x_1,...,X_k = x_k$ $(1 \leq k \leq n)$, is denoted as $\varphi_k(x_1,...,x_k)$ and its associated minimum posterior expected loss as $V_k(x_1,...,x_k)$. If no observation has been taken yet, $\varphi_0(x_0)$ and $V_0(x_0)$ denote the action $a_1(x_0)$, $a_2(x_0)$, or $a_3(x_0)$ which yields the smallest prior expected loss and its associated minimum prior expected loss, respectively.

Next, $\varphi_{n-1}(x_1,...,x_{n-1})$ and $V_{n-1}(x_1,...,x_{n-1})$ are computed at stage (n-1) of sampling. At this stage of sampling, however, we must also take into account the possible action of continuing sampling, $a_4(x_1,...,x_k)$. Hence, $V_{n-1}(x_1,...,x_{n-1})$ at stage (n-1) must be compared with the posterior expected loss associated with continuing sampling. At stage (n-1) of sampling, the posterior expected loss associated with taking one more observation, $E[V_n(x_1,...,x_{n-1},X_n) \mid X_1 = x_1,...,X_{n-1} = x_{n-1}]$, is computed as follows:

$$E[V_n(x_1,...,x_{n-1},X_n) \mid X_1 = x_1,...,X_{n-1} = x_{n-1}] =$$

$$\sum_{x_n=0}^{x_n=1} V_n(x_1,...,x_n) * P(X_n \mid X_1 = x_1,...,X_{n-1} = x_{n-1}), \quad (15)$$

where $P(X_n \mid X_1 = x_1,...,X_{n-1} = x_{n-1})$ denotes the conditional distribution of X_n, given the observed item response vector $(x_1,...,x_{n-1})$. This is also called the posterior predictive distribution of X_n at stage (n-1) of sampling. In the next section it will be indicated how, generally, the posterior predictive distribution of X_k $(1 \le k \le n)$, given the observed item response vector $(x_1,...,x_{k-1})$, can be computed. Note that (15) averages the posterior expected loss associated with each of the possible future decison outcomes relative to the probability of observing those outcomes (Lewis & Sheehan, 1990).

Following Lewis and Sheehan (1990), the minimum conditional Bayes risk at stage (n-1) of sampling, given $X_1 = x_1,...,X_{n-1} = x_{n-1}$, is defined as:

$$R_{n-1}(x_1,...,x_{n-1}) = \min\{V_{n-1}(x_1,...,x_{n-1}), E[V_n(x_1,...,x_{n-1},X_n) \mid X_1 = x_1,...,X_{n-1} = x_{n-1}]\}. \quad (16)$$

Let the optimal rule for the sequential four-action mastery problem at stage (k-1) $(1 \le k \le n)$, given $X_1 = x_1,...,X_{k-1} = x_{k-1}$, be defined as $d_{k-1}(x_1,...,x_{k-1})$, where $d_0(x_0)$ denotes the decision whether or not to take at least one observation. Then, $d_{n-1}(x_1,...,x_{n-1})$ can be obtained by comparing $V_{n-1}(x_1,...,x_{n-1})$ and $E[V_n(x_1,...,x_{n-1},X_n) \mid X_1 = x_1,..., X_{n-1} = x_{n-1}]$ with each other. Hence, it follows:

$$d_{n-1}(x_1,...,x_{n-1}) = \quad (17)$$

$$\begin{cases} \varphi_{n-1}(x_1,...,x_{n-1}) & \text{if } R_{n-1}(x_1,...,x_{n-1}) = V_{n-1}((x_1,...,x_{n-1}) \\ \text{continue sampling} & \text{if } R_{n-1}(x_1,...,x_{n-1}) = E[V_n((x_1,...,x_{n-1},X_n) \mid X_1 = x_1,...,X_{n-1} = x_{n-1}]. \end{cases}$$

In the case of equality between $V_{n-1}(x_1,...,x_{n-1})$ and $E[V_n(x_1,...,x_{n-1},X_n) \mid X_1 = x_1,...,X_{n-1} = x_{n-1}]$ it does not matter whether or not the decision-maker takes one more observation.

Let the minimum conditional Bayes risk at stage n of sampling, given $X_1 = x_1,...,X_n = x_n$, be defined as $V_n(x_1,...,x_n)$. Then, generally, the minimum conditional Bayes risk at stage (k-1), given $X_1 = x_1,...,X_{k-1} = x_{k-1}$, is computed inductively as a function of the minimum conditional Bayes risk at stage k ($1 \leq k \leq n$) as follows:

$$R_{k-1}(x_1,...,x_{k-1}) = \min\{V_{k-1}(x_1,...,x_{k-1}), E[R_k(x_1,...,x_{k-1},X_k) \mid X_1 = x_1,...,X_{k-1} = x_{k-1}]\}, \quad \textbf{(18)}$$

where the posterior expected loss associated with taking one more observation at stage (k-1) of sampling, $E[R_k(x_1,...,x_{k-1},X_k) \mid X_1 = x_1,...,X_{k-1} = x_{k-1}]$, is computed as follows ($1 \leq k \leq n$):

$$E[R_k(x_1,...,x_{k-1},X_k) \mid X_1 = x_1,...,X_{k-1} = x_{k-1}] =$$

$$\sum_{x_k=0}^{x_k=1} R_k(x_1,...,x_k) * P(X_k \mid X_1 = x_1,...,X_{k-1} = x_{k-1}). \quad \textbf{(19)}$$

Analogous to the computation at stage (n-1), $\varphi_{n-2}(x_1,...,x_{n-2})$ and $V_{n-2}(x_1,...,x_{n-2})$ are now computed at stage (n-2) of sampling. Next, using (18)-(19), $E[R_{n-1}(x_1,...,x_{n-2},X_{n-1}) \mid X_1 = x_1,...,X_{n-2} = x_{n-2}]$ and $R_{n-2}(x_1,...,x_{n-2})$ are computed at stage (n-2). Finally, analogous to the computation of $d_{n-1}(x_1,...,x_{n-1})$, $d_{n-2}(x_1,...,x_{n-2})$ is computed at stage (n-2) by comparing $V_{n-2}(x_1,...,x_{n-2})$ and $E[R_{n-1}(x_1,...,x_{n-2},X_{n-1}) \mid X_1 = x_1,...,X_{n-2} = x_{n-2}]$ with each other. Following the same computational backward scheme, $d_{n-3}(x_1,...,x_{n-3}),...,d_0(x_0)$ are computed.

Computation of Posterior Predictive Distribution

In this section, it will be indicated how the posterior predictive distribution $P(X_k \mid X_1 = x_1,...,X_{k-1} = x_{k-1})$ in (19) can be computed ($1 \leq k \leq n$). From Bayes' theorem, it follows that:

$$P(X_k \mid X_1 = x_1,...,X_{k-1} = x_{k-1}) = P(X_1 = x_1,...,X_k = x_k)/P(X_1 = x_1,...,X_{k-1} = x_{k-1}). \quad \textbf{(20)}$$

Since the binomial model was adopted for the psychometric model involved, it follows from (2) that

$$P(X_1 = x_1,\ldots,X_k = x_k|t) = t^{s_k}(1-t)^{k-s_k}. \quad (21)$$

Furthermore, the p.d.f. of T was assumed to be distributed according to a beta distribution $B(\alpha,\beta)$ with parameters α and β ($\alpha, \beta > 0$) in the standard interval [0,1]:

$$p(t) = [\Gamma(\alpha+\beta)t^{\alpha-1}(1-t)^{\beta-1}]/[\Gamma(\alpha)\Gamma(\beta)], \quad (22)$$

where Γ is the usual gamma function.

Multiplying (21) and (22) and integrating out t yields the unconditional distribution of $(X_1,\ldots,X_k)$:

$$P(X_1 = x_1,\ldots,X_k = x_k) =$$
$$[\Gamma(\alpha+\beta)\Gamma(\alpha+s_k)\Gamma(\beta+k-s_k)]/[\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+k)]. \quad (23)$$

Similarly, the unconditional distribution of $(X_1,\ldots,X_{k-1})$ is equal to:

$$P(X_1 = x_1,\ldots,X_{k-1} = x_{k-1}) =$$
$$[\Gamma(\alpha+\beta)\Gamma(\alpha+s_{k-1})\Gamma(\beta+k-1-s_{k-1})]/[\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+k-1)]. \quad (24)$$

It now follows from (20), (23), and (24) that the posterior predictive distribution of X_k, given the observed item response vector $(x_1,\ldots,x_{k-1})$, can be written as:

$$P(X_k \mid X_1 = x_1,\ldots,X_{k-1} = x_{k-1}) =$$
$$[\Gamma(\alpha+s_k)\Gamma(\beta+k-s_k)\Gamma(\alpha+\beta+k-1)]/[\Gamma(\alpha+s_{k-1})\Gamma(\beta+k-1-s_{k-1})\Gamma(\alpha+\beta+k)]. \quad (25)$$

Since $s_k = s_{k-1}$ and $s_k = s_{k-1}+1$ for $x_k = 0$ and 1, respectively, and using the well-known identity $\Gamma(j+1) = j\Gamma(j)$, it finally follows from (25) that:

$$P(X_k \mid X_1 = x_1,\ldots,X_{k-1} = x_{k-1}) = \begin{cases} (\beta - s_{k-1} + k - 1)/(\alpha+\beta+k-1) & \text{if } x_k = 0 \\ (\alpha + s_{k-1})/(\alpha+\beta+k-1) & \text{if } x_k = 1. \end{cases} \quad (26)$$

An Empirical Example

The procedures for computing the optimal sequential number of items were applied to a computerized four-action mastery test for concept-learning in medicine for freshmen. Concept-learning is the process in which subjects learn to categorize objects, processes or events, for instance, formation of diagnostic skills in medicine or psychology (see Tennyson and Cocchiarella, 1986, for a complete review of the theory of concept-learning).

Information from the group to which the student belongs was available in the form of data from a pretest for a sample of 76 freshmen in a medical program. The pretest consisted of 30 multiple-choice items and had possible test scores ranging from 0-30. The mean and KR-21 reliability coefficient were estimated as 16 and 0.81, respectively. Hence, it follows from (9) that α and β were estimated as 3.75 and 3.28, respectively.

The fit of the pretest data to the binomial model with the assumed beta distribution for prior true level of functioning was checked by comparing the theoretical score distribution with the empirical observed score distribution. Keats and Lord (1962) have shown that the theoretical score distribution is the negative hypergeometric distribution. The results of the chi-square test showed a satisfactory fit at a significance level of 0.05.

The instructors of the program considered students as having mastered the present concept successfully if they had mastered at least 60% of the total number of items covering the subject matter of that concept (i.e., true mastery). Therefore, t_{c2} was fixed at 0.6. Furthermore, nonmastery was declared if students had mastered less than 50% of the total number of items covering the subject matter of the present concept (i.e., true nonmastery). Therefore, t_{c1} was fixed at 0.5.

Finally, the constant cost for administering one random item was assumed to be rather small. Therefore, the value of e was set equal to 0.01.

Results with Linear Loss and a Beta Prior for T

First, the case of linear loss and a beta distribution for prior knowledge about T is considered. Taking into account the requirements discussed earlier, the loss parameters were empirically assessed by the instructors of the program yielding the following result: $b_1 = 8$, $b_2 = 3$, and $b_3 = 1$. For these values of the loss parameters, the right-hand side of (8) turned out to be equal to 0.44.

The appropriate action (i.e., nonmastery, partial mastery, mastery, or continue sampling) is depicted in Table 2 as a closed interval for a maximum of 30 items (i.e., n = 30) at each stage of sampling k ($0 \leq k \leq n$) for different number correct score s_k ($0 \leq s_k \leq k$).

Table 2 has been constructed by applying the following backward induction computational scheme. First, the appropriate action and its associated minimum posterior expected loss at the final stage of sampling have been determined; that is, $\varphi_{30}(x_1,...,x_{30})$ and $V_{30}(x_1,...,x_{30})$, have been computed for $s_{30} = 0,...,30$. More specifically, nonmastery was declared for those values of s_{30} for which $E(T \mid s_{30}) \leq 0.44$, partial mastery was declared for those values of s_{30} for which $0.44 < E(T \mid s_{30}) < 0.6$, and mastery was declared for those values of s_{30} for which $E(T \mid s_{30}) \geq 0.6$. Note that it can be inferred from Table 2 that the cutting scores $s_{c1}(30)$ and $s_{c2}(30)$ are equal to 12 and 19, respectively.

Similarly, the appropriate action nonmastery, partial mastery, or mastery and its associated minimum posterior expected loss have been computed after 29 items for $s_{29} = 0,...,29$ (i.e., $\varphi_{29}(x_1,...,x_{29})$ and $V_{29}(x_1,...,x_{29})$). Next, using (18), (19), (26), and the minimum posterior expected losses calculated at the final stage of sampling, the posterior expected loss associated with taking one more observation at stage 29 of sampling is computed for $s_{29} = 0,...,29$ (i.e., $E[V_{30}(x_1,...,x_{29},X_{30}) \mid X_1 = x_1,...,X_{29} = x_{29}]$). These values are compared to the minimum posterior expected losses after stopping after 29 items in order to compute the conditional Bayes risk at stage 29 of sampling. Using (17), the appropriate action nonmastery, partial mastery, mastery, or continue sampling is determined at stage 29. Similarly, the appropriate action is determined at stage 28 until stage 0 of sampling. A computer program called LINEAR was developed to determine the appropriate action at each stage of sampling. A copy of the program LINEAR is available from the author upon request.

As can be seen from Table 2, regardless of the observed number-correct score s_k, the decision-maker takes at least five observations. Furthermore, Table 2 shows that a student whose posterior expectation of T is in the region of the intersection of the loss lines $l(a_1(x_1,...,x_k),t)$ and $l(a_2(x_1,...,x_k),t)$ or in the intersection of the loss lines $l(a_2(x_1,...,x_k),t)$ and $l(a_3(x_1,...,x_k),t)$, it is hard to classify him /her as a nonmaster, partial master, or master. Hence, longer tests are needed for such students. On the other side, shorter tests can be provided for students whose posterior expectation of T is not in the region of the intersection of these loss lines.

Table 2. Appropriate action calculated by stage of sampling and number-correct in case of linear loss.

Stage of sampling	Appropriate Action by Number-Correct				
	Nonmastery	Continue	Partial Mastery	Continue	Mastery
0		0			
1		[0,1]			
2		[0,2]			
3		[0,3]			
4		[0,4]			
5		[0,4]			5
6	0	[1,5]			6
7	0	[1,5]			[6,7]
8	0	[1,6]			[7,8]
9	[0,1]	[2,7]			[9,0]
10	[0,1]	[2,7]			[8,10]
11	[0,2]	[3,8]			[9,11]
12	[0,2]	[3,8]			[9,12]
13	[0,3]	[4,6]	7	[8,9]	[10,13]
14	[0,3]	[4,9]			[10,14]
15	[0,4]	[5,7]	8	[9,10]	[11,15]
16	[0,4]	[5,7]	[8,9]	[10,11]	[12,16]
17	[0,5]	[6,8]	9	[10,11]	[12,17]
18	[0,5]	[6,8]	[9,10]	[11,12]	[13,18]
19	[0,6]	[7,9]	10	[11,12]	[13,19]
20	[0,6]	[7,9]	[10,11]	[12,13]	[14,20]
21	[0,7]	[8,9]	[10,12]	13	[14,22]
22	[0,7]	[8,10]	[11,12]	[13,14]	[15,22]
23	[0,8]	[9,10]	[11,13]	[14,15]	[16,23]
24	[0,8]	[9,11]	[12,14]	15	[16,24]
25	[0,9]	[10,11]	[12,14]	[15,16]	[17,25]
26	[0,9]	[10,11]	[12,15]	16	[17,26]
27	[0,10]	[11,12]	[13,15]	[16,17]	[18,27]
28	[0,10]	[11,12]	[13,16]	17	[18,28]
29	[0,11]	12	[13,17]	18	[19,29]
30	[0,12]		[13,18]		[19,30]

Finally, it can be inferred from Table 2 that with increasing number of items being administered, the chances of being classified as a nonmaster, partial master, or master increases.

Let us assume that the sequential decision procedure starts with administering one ramdomly selected item and stops after declaring nonmastery, partial mastery, or mastery. Hence, the sequential decision procedure proceeds only after the continue sampling decision. Then, it can easily be inferred from Table 2 that the optimal sequential rule can be depicted in

Table 3 at each stage of sampling k ($1 \leq k \leq 30$) for different number-correct score s_k ($0 \leq s_k \leq k$) as follows:

Table 3. Optimal sequential rule calculated by stage of sampling and number-correct in case of linear loss.

Stage of sampling	Optimal Sequential Rule by Number-Correct				
	Nonmastery	Continue	Partial Mastery	Continue	Mastery
1		[0,1]			
2		[0,2]			
3		[0,3]			
4		[0,4]			
5		[0,4]			5
6	0	[1,5]			
7		[1,5]			6
8		[1,6]			
9	1	[2,7]			
10		[2,7]			8
11	2	[3,8]			
12		[3,8]			9
13	3	[4,6]	7	[8,9]	
14		[4,9]			10
15	4	[5,7]	8	[9,10]	
16		[5,7]	[8,9]	[10,11]	
17	5	[6,8]		[10,11]	12
18		[6,8]	[9,10]	[11,12]	
19	6	[7,9]		[11,12]	13
20		[7,9]	[10,11]	[12,13]	
21	7	[8,9]	10 or 12	13	14
22		[8,10]		[13,14]	
23	8	[9,10]	11 or 13	[14,15]	
24		[9,11]	14	15	16
25	9	[10,11]	12	[15,16]	
26		[10,11]	12 or 15	16	17
27	10	[11,12]		[16,17]	
28		[11,12]	13 or 16	17	18
29	11	12	13 or 17	18	
30	12		13 or 18		19

Note that not all possible number-correct scores s_k are necessarily present at each stage of sampling k, because it is assumed in Table 3 that the optimal sequential rule stops after declaring nonmastery, partial mastery, or mastery. For instance, the number-correct score s_6

BEST COPY AVAILABLE

can only take the values 0 until 5, and thus, not the value 6. This is because mastery was declared for $s_5 = 5$, implying the optimal sequential rule stops for this value of s_5.

Results with Threshold Loss and a Beta Prior for T

Next, the case of threshold loss and a beta prior for T is considered. Taking into account the requirements discussed earlier and assuming equal losses for the correct decisions l_{11}, l_{22}, and l_{33}, the losses from Table 1 were empirically assessed by the instructors of the program yielding the following result:

Table 4. Threshold loss table at stage k ($1 \leq k \leq n$) of sampling for empirical example.

True Level / Action	$T \leq t_{c1}$	$t_{c1} < T < t_{c2}$	$T \geq t_{c1}$
$a_1(x_1, \ldots, x_k)$	ke	4 + ke	7 + ke
$a_2(x_1, \ldots, x_k)$	1 + ke	ke	2 + ke
$a_3(x_1, \ldots, x_k)$	3 + ke	1 + ke	ke

Note that l_{23} was assessed larger than l_{21} for this specific empirical example. Using the numerical values for the loss parameters l_{ij} (i,j = 1,2,3) of Table 4, the appropriate action nonmastery, partial mastery, mastery, or continue sampling is depicted in Table 5 for a maximum of 30 items at each stage of sampling k ($0 \leq k \leq n$) for different number correct score s_k ($0 \leq s_k \leq k$) as a closed interval again.

Table 5 was constructed by using the same backward induction computational scheme as in the construction of Table 2. Doing so, the appropriate action at each stage of sampling k ($0 \leq k \leq n$) for the fixed-length three-action mastery problem (i.e., $\varphi_k(x_1,\ldots,x_k)$) was determined by examining if the inequalities in (12) and (14) were satisfied. More specifically, nonmastery was declared for those values of s_k ($0 \leq s_k \leq k$) for which the left-hand sides of (14) were equal to or smaller than zero, partial mastery was declared for those values of s_k for which the left-hand sides of (14) and (12) were larger and smaller than zero, respectively, and mastery was declared for those values of s_k for which the left-hand sides of (12) were equal to or larger than zero. Using numerical procedures for calculating the cumulative posterior distributions $P(T \geq t_{c1} \mid s_k)$ and $P(T \geq t_{c2} \mid s_k)$, a computer program called THRESHOLD was

developed to determine the appropriate action. A copy of the program THRESHOLD is available from the author upon request.

Table 5. Appropriate action calculated by stage of sampling and number-correct in case of linear loss.

Stage of sampling	Appropriate Action by Number-Correct				
	Nonmastery	Continue	Partial Mastery	Continue	Mastery
0		0			
1		[0,1]			
2		[0,2]			
3		[0,3]			
4		[0,4]			
5		[0,5]			
6		[0,5]			6
7		[0,6]			7
8		[0,7]			8
9	0	[1,7]			[8,9]
10	0	[1,8]			[9,10]
11	0	[1,8]			[9,11]
12	[0,1]	[2,9]			[10,12]
13	[0,1]	[2,10]			[11,13]
14	[0,2]	[3,10]			[11,14]
15	[0,2]	[3,11]			[12,15]
16	[0,3]	[4,11]			[12,16]
17	[0,3]	[4,12]			[13,17]
18	[0,4]	[5,12]			[13,18]
19	[0,4]	[5,13]			[14,19]
20	[0,5]	[6,13]			[14,20]
21	[0,5]	[6,14]			[15,21]
22	[0,6]	[7,14]			[15,22]
23	[0,6]	[7,15]			[16,23]
24	[0,7]	[8,11]	12	[13,15]	[16,24]
25	[0,8]	[9,11]	[12,13]	[14,16]	[17,25]
26	[0,8]	[9,11]	[12,13]	[14,16]	[17,26]
27	[0,9]	[10,12]	[13,14]	[15,17]	[18,27]
28	[0,10]	[11,12]	[13,15]	[16,17]	[18,28]
29	[0,11]	12	[13,16]	17	[18,29]
30	[0,12]		[13,17]		[18,30]

As can be seen from Table 5, analogous to the situation with linear loss, the decision-maker takes at least five observations. Furthermore, Table 5 shows that continue sampling decisions in the region between the actions partial mastery and mastery are taken for the first time after 23 items have been administered. Continue sampling decisions in the region

between the actions nonmastery and partial mastery, however, are taken already after 8 items have been administered.

A possible explanation for this finding might be that the losses associated with taking false nonmastery decisions are rather large relative to the losses associated with taking false partial mastery decisions (i.e., 4 and 7 relative to 1 and 2), whereas the losses associated with taking false partial mastery and mastery decisions (i.e., 1 and 2 relative to 3 and 1) do not

Table 6. Optimal sequential rule calculated by stage of sampling and number-correct in case of threshold loss.

Stage of sampling	Optimal Sequential Rule by Number-Correct				
	Nonmastery	Continue	Partial Mastery	Continue	Mastery
1		[0,1]			
2		[0,2]			
3		[0,3]			
4		[0,4]			
5		[0,5]			
6		[0,5]			6
7		[0,6]			
8		[0,7]			
9	0	[1,7]			8
10		[1,8]			
11		[1,8]			9
12	1	[2,9]			
13		[2,10]			
14	2	[3,10]			11
15		[3,11]			
16	3	[4,11]			12
17		[4,12]			
18	4	[5,12]			13
19		[5,13]			
20	5	[6,13]			14
21		[6,14]			
22	6	[7,14]			15
23		[7,15]			
24	7	[8,11]	12	[13,15]	16
25	8	[9,11]	[12,13]	[14,16]	
26		[9,11]	12	[14,16]	17
27	9	[10,12]	14	[15,17]	
28	10	[11,12]	13 or 15	[16,17]	18
29	11	12	13 or 16	17	18
30	12		13 or 17		18

differ that much. Consequently, it seems better to continue sampling in the region between the actions nonmastery and partial mastery in order to avoid relatively large losses associated with taking false decisions.

Finally, analogous to the construction of Table 3 from Table 2, the optimal sequential rule can be inferred from Table 5 at each stage of sampling k ($1 \leq k \leq 30$) and for different number-correct score s_k ($0 \leq s_k \leq k$) again. The result is depicted in Table 6.

Conclusions and Some New Lines of Research

In this paper, using the framework of empirical Bayesian decision theory, optimal sequential rules for the four-action mastery problem (nonmastery, partial mastery, mastery, and continuing sampling) were derived. The procedures were demonstrated by an empirical example for concept learning in medicine. Both for threshold and linear loss, optimal sequential rules were derived with prior knowledge assumed to be represented by a beta distribution.

The results indicated that, regardless of the observed number-correct score, the decision-maker takes at least five observations for both loss structures. Furthermore, it turned out that the chances of being classified as a nonmaster, partial master, or master increased if the number of items administered increased. This result was in accordance with our expectations.

There are a few new lines of research arising from the application of (empirical) Bayesian decision theory to sequential mastery testing. The first is the extension of determining the optimal sequential decision rules to the case that, in addition to the actions nonmastery, partial mastery, mastery, and administer randomly one more item, still another action is open to the decision-maker (e.g., mastery with distinction). Following the same line of reasoning as in the situation where there are four actions open to the decision-maker, the optimal sequential rules can easily be generalized to this sequential five-action mastery problem.

Two, it might be assumed that guessing and carelessness have to be taken into account. Morgan (1979) has developed a model with corrections for guessing and carelessness within a Bayesian decision-theoretic framework (see also van den Brink & Koele, 1980). The results of a computer simulation of the model indicate that guessing and carelessness may

BEST COPY AVAILABLE

markedly affect the determination of cutting scores, and hence the accuracy of the sequential decision procedures.

Third, it might also be assumed that no prior knowledge about true level of functioning is available. In these circumstances, the maximin procedure might be an appropriate framework (e.g., Huynh, 1980; Veldhuyzen, 1982), which requires no prior distribution regarding true level of functioning. As an aside, it might be noted that a maximin rule can be conceived as a rule that is based on minimization of posterior expected loss as well, but under the restriction that the prior is the least favorable of the class of priors (e.g., Ferguson, 1967, Sect. 1.6).

The last line is research into other prior distributions, psychometric models (e.g., standard-normal model), and loss structures than the ones assumed here. For example, the normal ogive function (Novick & Lindley, 1979) which takes loss to be a nonlinearly function of the true level of functioning, might be a realistic representation of the losses actually incurred. This loss function does not only have realistic properties but also can be combined nicely with a standard normal distribution for the psychometric model.

References

Angoff, W.H. (1971). Scales, norms and equivalent scores. In R.L. Thorndike (Ed.), *Educational Measurement*. Washington, D.C.: American Council on Education.

Bechtel, G.G. (1976). *Multidimensional preference scaling*. Den Haag: Mouton.

Ben-Shakhar, G., & Beller, M. (1983). An application of a decision-theoretic model to a quota-free selection problem. *Journal of Applied Psychology, 68*, 137-146.

Chuang, D.T., Chen, J.J., & Novick, M.R. (1981). Theory and practice for the use of cut-scores for personnel decisions. *Journal of Educational Statistics, 6*, 129-152.

Cronbach, L.J., & Gleser, G.L. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana IL: University of Illinois Press.

Davis, C.E., Hickman, J., & Novick, M.R. (1973). *A primer on decision analysis for individually prescribed instruction* (ACT Technical Bulletin No. 17). Iowa City, IA: The American College Testing Program.

DeGroot, M.H. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.

De Gruijter, D.N.M., & Hambleton, R.K. (1984). On problems encountered using decision theory to set cutoff scores. *Applied Psychological Measurement, 8*, 1-8.

Ebel, R.L. (1972). *Essentials of educational measurement*. New York: Prentice Hall.

Ferguson, R.L. (1969a). *Computer-assisted criterion-referenced measurement* (Working Paper No. 41). Pittsburgh PA: University of Pittsburgh Learning and Research Development Center. (Eric Documentation Reproduction No. ED 037 089).

Ferguson, R.L. (1969b). *The development, implementation, and evaluation of a computer-assisted branched test for a program of individually prescribed instruction*. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh PA.

Ferguson, T.S. (1967). *Mathematical statistics: A decision theoretic approach*. New York: Academic Press.

Gray, D.A. (1988). Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika, 53*, 383-392.

Hambleton, R.K., & Novick, M.R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement, 10*, 159-170.

Hogg, R.V., & Craig, A.T. (1978). *Introduction to mathematical statistics*. Toronto, Canada: Macmillan Co.

Huynh, H. (1976). Statistical considerations of mastery scores. *Psychometrika, 41*, 65-79.

Huynh, H. (1980). A nonrandomized minimax solution for passing scores in the binomial error model. *Psychometrika, 45*, 167-182.

Johnson, N.L., & Kotz, S. (1970). *Distributions in statistics: Continuous univariate distributions* -2. Boston: Houghton Miffin.

Keats, J.A., & Lord, F.M. (1962). A theoretical distribution for mental test scores. *Psychometrika, 27*, 59-72.

Kingsbury, G.G., & Weiss, D.J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In D.J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 257-283). New York: Academic Press.

Lehmann, E.L. (1959). Testing statistical theory (3rd ed.). New York: Macmillan.

Lewis, C., & Sheehan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement, 14*, 367-386.

Lindgren, B.W. (1976). *Statistical theory* (3rd ed.). New York: Macmillan.

Luce, R.D., & Raiffa, H. (1957). *Games and decisions*. New York: John Wiley and Sons.

Morgan, G. (1979). *A criterion-referenced measurement model with corrections for guessing and carelessness* (Occasional Paper No. 13). Victoria: The Australian Council for Educaional Research Limited.

Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14, 3-19.

Novick, M.R., & Jackson, P.H. (1974). *Statistical methods for educational and psychological research*. New York: McGraw-Hill.

Novick, M.R., & Lewis, C. (1974). Prescribing test length for criterion-referenced measurement. In C.W. Harris, M.C. Alkin, & W.J. Popham (Eds.), *Problems in criterion-referenced measurement*. CSE monograph series in evaluation, No. 3 Los Angeles: Center for the Study of Evaluation, University of California.

Novick, M.R., & Lindley, D.V. (1979). The use of more realistic utility functions in educational applications. *Journal of Educational Measurement, 15*, 181-191.

Pearson, K. (1930). *Tables for statisticians and biometricians*. London: Cambridge University Press.

Raju, N.S., Steinhaus, S.D., Edwards, J.E., & Lehessio, J. (1991). A logistic regression model for personnel selection. *Applied Psychological Measurement, 15*, 139-152.

Reckase, M.D. (1983). A procedure for decision making using tailored testing. In D.J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 237-257). New York: Academic Press.

Sheehan, K., & Lewis, C. (1992). Computerized mastery testing with nonequivalent testlets. *Applied Psychological Measurement, 16*, 65-76.

Spray, J.A., & Reckase, M.D. (1996). Comparison of SPRT and sequential Bayes procedures for classifying examinees into tow categories using a computerized test. *Journal of Educational and Behavioral Statistics, 21*, 405-414.

Swaminathan, H., Hambleton, R.K., & Algina, J. (1975). A Bayesian decision theoretic procedure for use with criterion-referenced tests. *Journal of Educational Measurement, 12*, 87-98.

Tennyson, R.D., & Cocchiarella, M.J. (1986). An empirically based instructional design theory for teaching concepts. *Review of Educational Research, 56*, 40-71.

Torgerson, W.S. (1958). *Theory and methods of scaling*. New York: John Wiley and Sons.

van den Brink, W.P., & Koele, P. (1980). Item sampling, guessing and decision-making in achievement testing. *British Journal of Mathematical and Statistical Psychology, 33*, 104-108.

van der Gaag, N. (1990). *Empirische utiliteiten voor psychometrische beslissingen [Empirical utilities for psychometric decisions]*. Unpublished doctoral dissertation, University of Amsterdam, The Netherlands.

van der Gaag, N.L., Mellenbergh, G.J., & van den Brink, W.P. (1988). Empirical utility functions for pass/fail situations. *Methodika, 2*, 40-52.

van der Linden, W.J. (1980). Decision models for use with criterion-referenced tests. *Applied Psychological Measurement, 4*, 469-492.

van der Linden, W.J. (1990). Applications of decision theory to test-based decision making. In R.K. Hambleton & J.N. Zaal (Eds.), *New developments in testing: Theory and applications*, 129-155. Boston: Kluwer.

van der Linden, W.J., & Mellenbergh, G.J. (1977). Optimal cutting scores using a linear loss function. *Applied Psychological Measurement, 1*, 593-599.

van der Linden, W.J., & Vos, H.J. (1996). A compensatory approach to optimal selection with mastery scores. *Psychometrika, 61*, 155-172.

Veldhuijzen, N.H. (1982). Setting cutting scores: A minimum information approach. In W.J. van der Linden (Ed.), Aspects of criterion-referenced measurement. *Evaluation in Education: An International Review Series, 5*, 141-148.

Vos, H.J. (1990). Simultaneous optimization of decisions using a linear utility function. *Journal of Educational Statistics, 15*, 309-340.

Vos, H.J. (1991). Simultaneous optimization of the aptitude treatment interaction decision problem with mastery scores. In M. Wilson (Ed.), *Objective measurement: Theory into practice, 1*, 313-331. Norwood, New Yersey: Ablex Publishing Corporation.

Vos, H.J. (1995). A compensatory model for simultaneously setting cutting scores for selection-placement-mastery decisions. In I. Partchev (Ed.), *Multivariate analysis in the behavioral sciences: Philosophic to technical* (pp. 75-90). Sofia, Bulgaria: Academic Publishing House.

Vos, H.J. (1997). Simultaneous optimization of quota-restricted selection decisions with mastery scores. *British Journal of Mathematical and Statistical Psychology, 50*, 105-125.

Vos, H.J. (1998). A simultaneous approach to optimizing treatment assignments with mastery scores. *Multivariate Behavioral Research*, to appear.

Vrijhof, B.J., Mellenbergh, G.J., & van den Brink, W.P. (1983). Assessing and studying utility functions in psychometric decision theory. *Applied Psychological Measurement, 7*(3), 341-357.

Acknowledgments

The author is indebted to Wim J. van der Linden for his valuable comments and to Jan Gulmans for providing the data for the empirical example.

Titles of Recent Research Reports from the Department of Educational Measurement and Data Analysis. University of Twente, Enschede, The Netherlands.

RR-97-06 H.J. Vos, *Applications of Bayesian Decision Theory to Sequential Mastery Testing*

RR-97-05 W.J. van der Linden & Richard M. Luecht, *Observed-Score Equating as a Test Assembly Problem*

RR-97-04 W.J. van der Linden & J.J. Adema, *Simultaneous Assembly of Multiple Test Forms*

RR-97-03 W.J. van der Linden, *Multidimensional Adaptive Yesting with a Minimum Error-Variance Criterion*

RR-97-02 W.J. van der Linden, *A Procedure for Empirical Initialization of Adaptive Testing Algorithms*

RR-97-01 W.J. van der Linden & Lynda M. Reese, *A Model for Optimal Constrained Adaptive Testing*

RR-96-04 C.A.W. Glas & A.A. Béguin, *Appropriateness of IRT Observed Score Equating*

RR-96-03 C.A.W. Glas, *Testing the Generalized Partial Credit Model*

RR-96-02 C.A.W. Glas, *Detection of Differential Item Functioning using Lagrange Multiplier Tests*

RR-96-01 W.J. van der Linden, *Bayesian Item Selection Criteria for Adaptive Testing*

RR-95-03 W.J. van der Linden, *Assembling Tests for the Measurement of Multiple Abilities*

RR-95-02 W.J. van der Linden, *Stochastic Order in Dichotomous Item Response Models for Fixed Tests, Adaptive Tests, or Multiple Abilities*

RR-95-01 W.J. van der Linden, *Some decision theory for course placement*

RR-94-17 H.J. Vos, *A compensatory model for simultaneously setting cutting scores for selection-placement-mastery decisions*

RR-94-16 H.J. Vos, *Applications of Bayesian decision theory to intelligent tutoring systems*

RR-94-15 H.J. Vos, *An intelligent tutoring system for classifying students into Instructional treatments with mastery scores*

RR-94-13 W.J.J. Veerkamp & M.P.F. Berger, *A simple and fast item selection procedure for adaptive testing*

RR-94-12 R.R. Meijer, *Nonparametric and group-based person-fit statistics: A validity study and an empirical example*

RR-94-10 W.J. van der Linden & M.A. Zwarts, *Robustness of judgments in evaluation research*

RR-94-9 L.M.W. Akkermans, *Monte Carlo estimation of the conditional Rasch model*

RR-94-8 R.R. Meijer & K. Sijtsma, *Detection of aberrant item score patterns: A review of recent developments*

RR-94-7 W.J. van der Linden & R.M. Luecht, *An optimization model for test assembly to match observed-score distributions*

RR-94-6 W.J.J. Veerkamp & M.P.F. Berger, *Some new item selection criteria for adaptive testing*

RR-94-5 R.R. Meijer, K. Sijtsma & I.W. Molenaar, *Reliability estimation for single dichotomous items*

RR-94-4 M.P.F. Berger & W.J.J. Veerkamp, *A review of selection methods for optimal design*

RR-94-3 W.J. van der Linden, *A conceptual analysis of standard setting in large-scale assessments*

RR-94-2 W.J. van der Linden & H.J. Vos, *A compensatory approach to optimal selection with mastery scores*

RR-94-1 R.R. Meijer, *The influence of the presence of deviant item score patterns on the power of a person-fit statistic*

RR-93-1 P. Westers & H. Kelderman, *Generalizations of the Solution-Error Response-Error Model*

RR-91-1 H. Kelderman, *Computing Maximum Likelihood Estimates of Loglinear Models from Marginal Sums with Special Attention to Loglinear Item Response Theory*

RR-90-8 M.P.F. Berger & D.L. Knol, *On the Assessment of Dimensionality in Multidimensional Item Response Theory Models*

RR-90-7 E. Boekkooi-Timminga, *A Method for Designing IRT-based Item Banks*

RR-90-6 J.J. Adema, *The Construction of Weakly Parallel Tests by Mathematical Programming*

RR-90-5 J.J. Adema, *A Revised Simplex Method for Test Construction Problems*

RR-90-4 J.J. Adema, *Methods and Models for the Construction of Weakly Parallel Tests*

RR-90-2 H. Tobi, *Item Response Theory at subject- and group-level*

RR-90-1 P. Westers & H. Kelderman, *Differential item functioning in multiple choice items*

Research Reports can be obtained at costs, Faculty of Educational Science and Technology, University of Twente, Mr. J.M.J. Nelissen, P.O. Box 217, 7500 AE Enschede, The Netherlands.